

BARBARA BATKO-TOKARZ
Uniwersytet Jagielloński
Kraków

Tradycja *Thesaurusa* P.M. Rogeta w leksykografii polskiej i europejskiej

Streszczenie

Celem tego artykułu jest przyjrzenie się polskiej i europejskiej tradycji badań i dzieł inspirowanych *Tezaurusem* P.M. Rogeta. Opisana zostanie po pierwsze, tradycja słowników tematycznych w Polsce wraz ze źródłami ich inspiracji (zwł. koncepcja Halliga i Wartburga), po drugie widoczna wyraźnie na takim tle słaba obecność dzieła Rogeta w leksykografii polskiej ze szczególnym uwzględnieniem słownika R. Zawilińskiego i powodów jego małej popularności. Następnie przedstawione zostaną przyczyny popularności koncepcji Rogetowskiej w Europie, a także głosy krytyczne wobec niej. W artykule zaprezentowane zostaną także nowsze przykłady wykorzystania koncepcji Rogeta ze szczególnym uwzględnieniem nowoczesnych komputerowych projektów leksykograficznych.

Słowa klucze: leksykografia, słownik pojęciowy, tezaursus

Tradition of P.M. Roget's *Thesaurus* in Polish and European lexicography

Abstract

The aim of this article is to look closely at Polish and European research traditions and reference works inspired by Roget's *Thesaurus*. The first part of the article presents the tradition of Polish thematic dictionaries and discusses their inspiration sources (especially the approach by Hallig and Wartburg). It will be claimed, proved and exemplified that Roget's work is not popular in Polish lexicography – reference will be made to the dictionary compiled by R. Zawiliński. The second part of the article discusses the reasons for the popularity of Roget's concept in Europe, as well as its criticism. The paper presents also the newest examples of the use of Roget's theory, in particular modern computerized lexicographic projects.

Key words: lexicography, topic dictionary, thesaurus

Thesaurus of English Words and Phrases P.M. Rogeta (1779–1869) wydany został w 1852 roku. Słownik ten zyskał duże uznanie czytelników i jest nieustannie obecny w leksykografii angielskojęzycznej. Dzieło to bardzo szybko pojawiło się w bibliotekach w prawie każdym domu wykształconych ludzi w Wielkiej Brytanii i Stanach Zjednoczonych, a następnie książka ta rozpowszechniła się na cały świat, o czym świadczy najdobitniej duża liczba reprintów i nowych edycji (Hüllen 2004: 1). Jak pisze T. Piotrowski,

tezaury, wzorowane na słynnym dziele Petera Marka Rogeta, to niemal instytucja w świecie anglosaskim. Typ słownika, zwanego tezaurem właśnie dzięki dziełu Rogeta, zrobił ogromną karierę w świecie współczesnym (Piotrowski 1994: 98).

Tezaurus Rogeta miał wiele różnych wersji i wydań zarówno w anglojęzycznej literaturze, jak również w innych językach. Szacuje się, że od pierwszego wydania sprzedano ok. 32 miliony egzemplarzy i było około 100 wznowień. W 1852 roku w tezaurucie znajdowało się 15 000 słów, obecnie jest ich 300 000. Tezaurus został więc wzbogacony o nowe słownictwo, choć układ znaczeń okazał się tak uniwersalny, że dopiero w latach 60. XX wieku zaistniała potrzeba jego zmodyfikowania, w celu umieszczenia nowo powstającego słownictwa np. technicznego i naukowego. Współcześnie materiał słownika jest dostępny także w Internecie (Żmigrodzki 2009: 178).

W leksykografii angielskojęzycznej, a także polskiej są prace, w których można znaleźć opis kolejnych wydań tezaurusa, a także licznych przeobrażeń i modyfikacji zarówno ilościowych, jak i jakościowych tego dzieła (Jarmasz, Szpakowicz 2001; Hüllen 2004, 2009; Piotrowski 1994). Dlatego bezzasadne byłoby powielanie tych informacji. Warto jednak podkreślić w tym miejscu ogromną różnorodność i wieloaspektowość tego materiału, powołując się w tym celu chociażby na ponad 20-stronicowy artykuł A. Klégra na temat tezaursów inspirowanych dziełem Rogeta. Sama ilość kryteriów leksykograficznych, na które czeski badacz podzielił tezaury wzorowane na dziele Rogeta, jest imponująca i wymowna:

1. regionalne odmiany w zależności od miejsca używania języka angielskiego (brytyjska, amerykańska, australijska),
2. organizacja formalna, czyli typ makrostruktury (tezaury tematyczne, przejściowe, alfabetyczne),
3. zastosowanie zasad onomazjologicznych lub semazjologicznych,
4. językowe lub encyklopedyczne preferencje tezaurusa,
5. perspektywa czasowa, jaką obejmuje się słownictwo (synchroniczna, diachroniczna),
6. cel wydania tezaurusa (dydaktyczny, naukowy, diagnostyczny),
7. zasięg lub aspekt języka, który jest wzięty pod uwagę (tezaury języka ogólnego, slangowe, terminologiczne, idiomatyczne),
8. złożoność jednostek leksykalnych (idiomy, cytaty, przysłowia),
9. ilość języków, które są używane (jednojęzyczny, dwujęzyczny, wielojęzyczny),

10. rodzaj nośnika (tezauryusy książkowe, elektroniczne, internetowe),
11. wiek użytkowników, do których skierowany jest tezaurus (dorośli, dzieci),
12. wielkość tezaury (pełna, kieszonkowa) (Klęgr 2000).

W tym artykule opisane zostaną tematy nieczęsto poruszane przez polskich badaczy, a więc po pierwsze, przyczyny słabego wykorzystania w leksykografii polskiej *Tezaurusa* Rogeta na tle polskiej tradycji słowników onomazjologicznych i źródeł ich inspiracji. Po drugie, uwaga poświęcona będzie powodom popularności Rogeta w leksykografii europejskiej, ale także głosom krytycznym wobec tej koncepcji. Omówione też zostanie wykorzystywanie tej koncepcji leksykograficznej w leksykografii europejskiej XXI wieku, szczególnie w nowo powstających słownikach i lingwistyce komputerowej.

Tradycja słowników niealfabetycznych w Polsce

Tradycja klasyfikacji słownictwa według kryterium tematycznego nie jest w Polsce zbyt rozwinięta, lecz niewątpliwie istnieje (Kajtoch 2008: 28) i w porównaniu z innymi krajami słowiańskimi jest stosunkowo popularna, na co wskazują choćby badania porównawcze słowników tematycznych w Polsce, Serbii i Chorwacji (Daković 2011).

Słowniki pojęciowe mają na świecie bardzo długą tradycję, stanowiąc podstawowy historycznie typ słowników, ponieważ w leksykografii europejskiej długo nie stosowano alfabetycznego układu wyrazów (Piotrowski 1994: 109). Także historia polskich słowników zaczyna się od słowników wielojęzycznych o układzie tematycznym, rzeczowym. Uporządkowanie niealfabetyczne pojawia się w niektórych fragmentach najstarszego słownika łacińsko-polskiego, tzw. *Wokabularza trydenckiego* z początku XV wieku (Urbańczyk 1962). Także w XVI-wiecznych słownikach Murmeliusza i Mymera zastosowano porządek rzeczowy (Gruszczyński 2000; Mikołajczak-Matysa 2005: 28; Sieradzka-Baziur 2011: 347). Stopniowo leksykografię polską zdominowały jednak słowniki o układzie alfabetycznym. Współczesna polska leksykografia onomazjologiczna reprezentowana jest w większości przez specjalistyczne słowniki (np. Anusiewicz, Skawiński 1996), brak ogólnego słownika tematycznego obejmującego całość słownictwa¹ (por. Sieradzka-Baziur 2011: 347). Powstają natomiast słowniki ogólne o charakterze dydaktycznym lub glottodydaktycznym, które są objętościowo niewielkie (np. Kita, Polański 2002). Niejednokrotnie też układ tematyczny pojawia się w takim słowniku jako pomocniczy (np. Zgółkowa 2013).

¹ Widoczne są jednak próby tworzenia takich słowników, np. w ramach powstającego *Wielkiego słownika języka polskiego PAN* hasła są klasyfikowane tematycznie. Do końca 2012 roku powstało 15 tysięcy haseł, prace planowane są do 2018 roku, co znacząco zwiększy ilość słownictwa. Opcja wyszukiwania za pomocą kryterium przynależności do określonego pola tematycznego powoduje, że słownik ten po zakończeniu prac będzie mógł funkcjonować także jako słownik tematyczny.

Warto w kontekście refleksji o słowniku Rogeta przyrzeć się źródłom inspiracji tego typu słowników w leksykografii polskiej. Według W. Kajtocha tworzą ją (2008: 28–45):

- 1) badania nad słownictwem tematycznym prowadzone w Krakowie przez W. Miodunkę i Z. Cygał-Krupę;
- 2) recepcja słownika F. Dornseiffa *Der deutsche Wortschatz nach Sachgruppen*;
- 3) recepcja projektu R. Halliga i W. von Wartburga;
- 4) tradycja związana z *Tezaurusem* M.P. Rogeta oraz jego polską wersją *Do-borem wyrazów* R. Zawilińskiego – której współcześnie w Polsce nie rozwinęto.

Najbardziej nośna okazała się w leksykografii polskiej recepcja projektu Halliga i Wartburga, dlatego od niej należy zacząć rozważania. Będzie ona doskonałym punktem odniesienia do refleksji nad polską recepcją *Thesaurusa* Rogeta.

Tradycja słownika Halliga i Wartburga w leksykografii polskiej

Na gruncie polskim zdecydowanie bardziej popularny niż tradycja *Tezaurusa* Rogeta okazał się podział R. Halliga i W. von Wartburga, którzy w 1952 roku stworzyli systematyczny słownik rzeczowy języka francuskiego i schemat strukturalny do opracowania podobnych słowników pt. *Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*. Został on bardzo szybko wykorzystany w polskich słownikach gwarowych z lat 50. i 60. o układzie rzeczowym (np. Kucała 1957). W trzydzieści lat po pracy M. Kucały schemat Halliga i Wartburga poważnie zmodyfikował A. Markowski (1990). Jak stwierdza W. Kajtoch:

nowsze prace leksykograficzne stosujące układ tematyczny, jeśli tylko nie traktują go swobodnie i bardzo ogólnie, na ogół trzymają się schematu Halliga i Wartburga – albo w wersji bliższej oryginałowi, albo bliżej modyfikacjom Markowskiego (Kajtoch 2008: 35).

Jednym z najnowszych pomysłów na wykorzystanie schematu Halliga i Wartburga jest projekt *Słownika pojęciowego języka staropolskiego* – nowoczesnego słownika komputerowego w sieci internetowej (od lipca 2011 r.). Rezultatem pracy zespołu będzie pierwszy w Polsce, pełny słownik języka polskiego w układzie pojęciowym prezentujący leksykę najstarszego etapu jego rozwoju, opracowany na materiale istniejącego alfabetycznego *Słownika staropolskiego* (Sieradzka-Baziur 2012).

Tradycja *Tezaurusa* Rogeta w leksykografii polskiej

Na zarysowanym powyżej tle wykorzystanie tradycji *Tezaurusa* M.P. Rogeta wypada szczególnie blado. W polskiej literaturze językoznawczej dzieło to nie jest dokładnie znane, choć znaleźli się badacze, którzy poświęcili mu uwagę (Miko-

lajczak-Matyja 2005; Miodunka 1989; Piotrowski 1994; Żmigrodzki 2009). Częstokroć jednak tezaurus ten był pomijany przez badaczy. Opisuje to T. Piotrowski (1994), wskazując na przykład, że choć A. Markowski (1990) omówił różne próby onomazjologicznego układu słownictwa w Polsce i w Europie, nie zamieścił w swojej książce nazwiska Rogeta (a także Zawilińskiego), choć odniesienia do Rogeta są bardzo częste w literaturze rosyjskiej, na którą on i inni językoznawcy polscy się powoływali. Brakuje także omówienia dzieła Rogeta w *Dziejach językoznawstwa w zarysie* A. Heinza (1978). Badacz ten podaje mnóstwo nazwisk czwarto- i pięciorzędnych lingwistów, których prace – według Piotrowskiego – do niczego się nie przysłużyły, a pomija koncepcję Rogeta, podając jedynie jego naśladowców, np. Dornseiffa (Piotrowski 1994: 101).

Niezauważenie tezaurusu Rogeta w Polsce budzi tym większe zdziwienie, że do niedawna Polska jako jedyny kraj słowiański mogła się pochwalić słownikiem ściśle wzorowanym na tezaurusie Rogeta, pt. *Dobór wyrazów. Słownik wyrazów bliskoznacznych i jednoznacznych. Do praktycznego użytku* R. Zawilińskiego (Kraków 1926 r.), które to dzieło jest wciąż największym słownikiem języka polskiego ułożonym w porządku onomazjologicznym. Autor, który wzorował się na wydaniu *Tezaurusu* z 1912 roku, za podstawę klasyfikacji świata pojęciowego utrwalonego w języku przyjął za Rogetem osiągnięcia psychologii klasycznej i przyrodoznawstwa. Jego słownik zawiera około 20 tys. jednostek leksykalnych. Są one jak w *Tezaurusie* Rogeta ułożone w sześciu podstawowych działach. W ramach tych działów wyodrębniono dokładnie 1000 grup tematycznych, zawierających słownictwo w ustalonym porządku (Piotrowski 1994: 99–100; Żmigrodzki 2009: 178).

Niepowodzenie *Doboru wyrazów* polegające na braku kolejnych wydań i prób naśladowania, a tym samym niepowodzenie koncepcji Rogetowskiej w Polsce, można wyjaśniać np. brakiem przygotowania polskich czytelników do odbioru takiego tekstu leksykograficznego (Miodunka 1989). T. Piotrowski zwraca też uwagę, że pojedyncze informacje na temat słownika Zawilińskiego i jego pierwowzoru nie zawsze są rzetelne, częstokroć odbiera się temu opracowaniu charakter naukowy, czemu sprzyjało powtarzane przez badaczy stwierdzenie Zawilińskiego ze wstępu, „że praktycznemu tedy użyciu ma służyć [ten słownik] (...) nie szuka tedy niczego, coby mu nadawało charakter naukowy” (Piotrowski 1994: 99). Kolejną przyczyną niezwracania uwagi na to opracowanie może być wynikające z braku precyzji nomenklatury leksykograficznej mylne utożsamianie *Tezaurusu* Rogeta i dzieła Zawilińskiego ze słownikiem synonimów, mimo że jest to słownik pojęciowy, w którym słownictwo zostało zgrupowane według ogólnego schematu, obejmującego całość ludzkiego doświadczenia (Mikołajczak-Matyja 2005: 76; Żmigrodzki 2009: 178). Wiąże to się zapewne z samym tytułem tego dzieła, które nie zostało nazwane tezaurusem, lecz *Słownikiem wyrazów bliskoznacznych i jednoznacznych*, a także wieloznacznością pojęcia „tezaurus” w lingwistyce polskiej, ale też w europejskiej i częstym mylnym utożsamianiem słowników pojęciowych ze słownikami synonimów (Klégr 2007: 7; Piotrowski 1994: 106).

Powody popularności koncepcji Rogeta w Europie

Na tak zarysowanym tle szczególnie wyraziście wypada bogata europejska tradycja słowników inspirowanych dziełem Rogeta. Warto zadać sobie pytanie, jakie są jej przyczyny. W. Hüllen zauważa, że szeroka dystrybucja tej książki, prawie porównywalna z Biblią, nie jest łatwa do jednoznacznego wyjaśnienia. Badacz ten wskazuje na dwa aspekty sukcesu: praktyczny – polegający na powszechnym używaniu tego słownika jako nowoczesnego narzędzia pomocnego w rozmowach oraz w pisaniu prac oraz teoretyczny – związany z założeniami lingwistycznymi słownika. Ostatecznie badacz stwierdza, że powodzenie tej pracy leży w szczególnie trafnym połączeniu teoretycznych i praktycznych aspektów tego dzieła (Hüllen 2004: 2). Jeśli chodzi o aspekt praktyczny, warto podkreślić cel używania tezauryusa, który niewątpliwie zdecydował o jego popularności. *Tezaurus* jako słownik pojęciowy przeznaczony był zwłaszcza dla użytkowników wyrafinowanych, odbiorców doskonale znających język, np. tłumaczy, pisarzy, autorów testów różnego typu. Nawet obecnie istnieje ogromna potrzeba takiego słownika w języku polskim (Żmigrodzki 2009: 180). Tym bardziej więc pojawienie się takiego dzieła w XIX wieku było wielce pożądane, ponieważ wspomagało pamięć i proces myślenia, pomagając w przekraczaniu ograniczeń językowych. Główną wartość swego dzieła Roget widział właśnie w ułatwieniu pracy piszącym. Dlatego nie interesowały go różnice wartości stylistycznych wyrazów, zamieszczał on bowiem w tezaursie wszystkie wyrazy, górnolotne, neutralne, stare, nowe, techniczne, literackie, by w ten sposób najlepiej naprowadzić użytkownika na właściwy trop przy tworzeniu nowego tekstu (Piotrowski 1994: 106–107). Atrakcyjność tego dzieła zwiększyła moda na krzyżówki w latach 20. i 30. XX wieku (Piotrowski 1994: 105).

Także teoretyczny pomysł na słownik stał się przyczyną popularności Rogeta w lingwistyce europejskiej. Zwracają na to uwagę badacze, wskazując dwa istotne aspekty leksykograficzne tego dzieła: system klasyfikacji pojęć oraz zasadę włączania do niego wyrazów (Piotrowski 1994: 105). Podkreślają oni połączenie w jedną całość dwóch znanych wcześniej typów słowników, czyli słownika synonimów ze słownikiem pojęciowym (Hüllen 2004: 2–3), choć utożsamianie dzieła Rogeta tylko ze słownikiem synonimów jest zdecydowanym nadużyciem (na ten temat obszernie: Piotrowski 1994: 106–107). W przeciwieństwie bowiem do słowników semazjologicznych Roget stworzył słownik onomazjologiczny, pojęciowy, w którym znaczenia nie są podporządkowane formom wyrazów, jak w słownikach alfabetycznych, ale formy wyrazów podporządkowane są znaczeniom. Roget wychodzi od pojęcia, które to klasyfikuje coraz dokładniej, aż dochodzi do grup wyrazów (Piotrowski 1994: 105). Drugim niebagatelnym aspektem tego dzieła jest uniwersalność systemu klasyfikacyjnego Rogeta. W. Hüllen (2009: 22–23) zwraca uwagę na to, że duża liczba edycji, które osiągnęły sukces, odzwierciedlała oryginalny schemat dzieła Rogeta, a ciągle wzrastający zbiór słów doskonale wpisuje się w dziewiętnastowieczną strukturę. Szczególnie brytyjskie wydania *Tezaurusa* nie zmieniały oryginalnego schematu, uważając go za uniwersalny i elastyczny (Piotrowski 1994: 112).

Z tych powodów dzieło P.M. Rogeta w leksykografii anglojęzycznej, ale także europejskiej stało się niekwestionowanym wzorem tezaursu rozumianego jako słownik pojęciowy (Burkhanov 2010: 245). Jest on przywoływany jako nadrzędne dzieło w tego typu słownictwie, nawet przez twórców, którzy nie korzystają z jego koncepcji. Dzieje się tak na przykład w hiszpańskim słowniku tematycznym (Moral 1998) klasyfikującym według nowego pomysłu ponad 60 000 słów i wyrażań współczesnego języka hiszpańskiego. Jego autor, opisując tradycję słowników tematycznych, na pierwszym miejscu wymienia właśnie słownik Rogeta, a dopiero później wzorowany na słowniku Rogeta *Diccionario ideológico* J. Casaresa, który jest pierwszym w leksykografii hiszpańskiej słownikiem tego typu (na temat tego słownika patrz: Miodunka 1989: 225–226). Co więcej, często autorzy we wstępie deklarują celowe odbieganie od wersji Rogeta, także i wtedy jest on jednak wyraźnym punktem odniesienia w leksykografii europejskiej, np. w teaurusie R. Jovanovicia *Sistematski rečnik srpskohrvatskoga jezika* wydanym w 1940 r. w Belgradzie (za: Daković 2011: 39).

Obecnie również powstają tezaursy wzorowane w sposób bezpośredni na słowniku Rogeta. Mowa tu o wydanym w 2007 r. teaurusie języka czeskiego (Klégr 2007) stworzonym na podstawie tłumaczenia jednej ze zmodyfikowanych wersji omawianego dzieła Rogeta (kieszonkowego tezaursu *The Penguin Pocket English Thesaurus*, 1985). Według autora czeskiego tezaursu ta wersja Rogetowskiego słownika, choć jest może mniej dokładna (mniejsza liczba kategorii), zostawia jednak więcej przestrzeni przy adaptacji klasyfikacji dla czeszczyzny (Klégr 2007: 12). Praca nad tym dziełem odbywała się w dwóch etapach. Pierwszy etap pracy polegał na roboczym przekładzie haseł z angielskiego tezaursu na czeski i tworzeniu elektronicznej bazy danych. Drugi etap był tworzeniem tezaursu, czyli sporządzaniem tekstu hasła na podstawie czeskich źródeł, a następnie dopasowaniem do tego tekstu roboczego przekładu, np. poprzez usuwanie elementów typowych dla angielszczyzny. W czeskim teaurusie zadbano nawet o możliwość porównania haseł z hasłami oryginału (Klégr 2007).

Głosy krytyczne wobec koncepcji Rogetowskiej

Nie brakuje jednak i głosów krytycznych wobec koncepcji Rogetowskiej, w których zwraca się uwagę zwłaszcza na to, że mimo iż dzieło to ciągle stoi na półkach w wielu domach anglojęzycznego świata, wcale nie jest ono wykorzystywane, ponieważ przestało być funkcjonalne. Przeminięła obecnie chociażby popularność Rogeta wynikająca z mody na krzyżówki w 1. poł. XX wieku. Choć ciągle na rynku amerykańskim i angielskim pojawiają się nowe wydania *Tezaursu* Rogeta, ale wydania te, np. wydanie specjalne z okazji 150. rocznicy *editio princeps* w 2002 r. (por. Hüllen 2009: 22–26), jak twierdzą niektórzy, są zwłaszcza okazją do zwiększenia coraz gorszych wyników sprzedaży tej książki. Już bowiem w latach 90. zauważono, że od dziesięciu lat nie zmieniała się znacząco liczba sprzedanych egzemplarzy (Piotrowski 1994: 105). Ponadto wielu boi się przyznać, że

w tym dziele mimo uzupełnień nowym słownictwem są jednak pewne braki (recenzja L. Norfolka w *The Guardian* <http://www.theguardian.com/books/2002/jul/13/referenceandlanguages.Highereducation1>).

Badacze wskazują też częstokroć, że szukanie hasła przez klasyfikację i system pojęć jest niefunkcjonalne i tylko teoretyczne, ponieważ to system nieintuicyjny i trudny dla użytkowników (m.in. Burkhanov 2010: 245). Układ tego słownika, niejako odtwarzający działanie pamięci, jest faktycznie niesłychanie skomplikowany (Kajtoch 2008: 39). Nawet w dziełach współczesnych, będących wierną kopią *Tezaurusa* Rogeta, można przeczytać, że choć są dwa sposoby wyszukiwania słownictwa, zalecane jest szukanie przez indeks, ponieważ szukanie przez system pojęć – pierwotnie jedyny, potem główny sposób szukania – w czasach współczesnych jest tylko teoretyczne przez swoją nieintuicyjność i skomplikowanie (Klégr 2007: 14). Niedogodności związane z wyszukiwaniem poszczególnych haseł były impulsem do zmieniania tradycyjnego układu Rogetowskiego, szczególnie w wydaniach amerykańskich, a zwłaszcza w *Roget International Thesaurus* (1992) (na temat tej wersji por. Jarmasz, Szpakowicz 2001: 119–120), ponieważ jak twierdzi jego red. R. Chapman, schemat Rogeta (platońsko-arystotelesowski, czyli abstrakcyjno-logiczny) zestarzał się kulturowo, a obecnie potrzebny jest porządek rozwojowo-egzystencjalny, rządzący się prawami skojarzenia i trwania. Trzymanie się natomiast starego porządku klasyfikacji powoduje sztuczne i niefunkcjonalne dla użytkownika wprowadzanie nowego słownictwa (np. komputer pod relacjami abstrakcyjnymi, bilet pod intelektem, żywność pod ruchem itp.) (za: Piotrowski 1994: 113). Jednym z najnowszych i chyba najbardziej wyrazistych przykładów rezygnacji z planu klasyfikacji Rogeta jest słownik *The Historical Thesaurus of the Oxford English*, który powstawał ok. 45 lat, a ostatecznie został wydany w dwóch tomach w 2009 roku i jest dostępny również w wersji elektronicznej za pośrednictwem Internetu (Esquibel 2011: 143–144). Jest to pierwszy historyczny tezaurus, ale także jednocześnie największy tezaurus na świecie. Jego twórcy budowali bazę danych metodą „bottom up”, gromadząc elementy leksykalne początkowo w ramach kategorii zaczerpniętych z klasycznego *Tezaurusa* Rogeta, ale szybko okazało się, że nie był on adekwatny i dostatecznie szczegółowy, aby sprostać takiemu wyzwaniu. Po niepowodzeniach związanych z wykorzystaniem koncepcji Rogeta autorzy zaczęli pracę na nowo, tworząc własny, nowy system kategorii tematycznych. Przeszli oni do tzw. folk taxonomy, która odwzorowuje sposób postrzegania świata przez przeciętnego człowieka (zwł. anglojęzycznego) (Esquibel 2011: 144; <http://public.oed.com/historical-thesaurus-of-the-oed/what-is-the-historical-thesaurus-of-the-oed/>; http://www.pw.org/content/world's_largest_thesaurus_published).

Wykorzystanie *Tezaurusa* Rogeta w informatyce

Opisując *Tezaurus* Rogeta, nie sposób pominąć jeszcze jednego aspektu. Słowo *tezaurus* przeżywa swój renesans dzięki technologii komputerowej, obecnie w informatyce czy naukach o informacji używane jest ono częstokroć. W kontekś-

cie wyszukiwania informacji nazwa *tezaurus* zaczęła być używana w końcu lat pięćdziesiątych XX w. (B. Sosińska-Kalata 2006: 152). Zjawisko to we wcześniejszych etapach rozwoju opisał już T. Piotrowski, zwracając uwagę na pierwsze próby przekładu maszynowego, jakie prowadzono w Cambridge Language Research Units (Piotrowski 1994: 115–116). Jak zauważają badacze,

uksztaltowaną w ciągu kilkudziesięciu lat koncepcję tezaurya stosowanego w wyszukiwaniu informacji zaliczyć można do największych i chyba najbardziej znanych osiągnięć w zakresie profesjonalnych metod i narzędzi działalności informacyjnej. Koncepcja ta nie tylko doczekała się standaryzacji w skali międzynarodowej, ale też wywarła duży wpływ na rozwój innych, historycznie wcześniejszych języków informacyjno-wyszukiwawczych, jak również wykorzystywana jest w kształtowaniu nowych narzędzi semantycznej organizacji i dostępu do cyfrowych zasobów informacyjnych (B. Sosińska-Kalata 2006: 150).

Omawiając to zagadnienie, warto zwrócić szczególną uwagę na inspirację tradycyjnymi tezaurusami w nowoczesnych projektach leksykograficznych wykorzystujących najnowszą technologię komputerową. Jak piszą badacze, twórcy tezaursów stosowanych w wyszukiwaniu informacji nawiązywali wyraźnie do pojęciowego sposobu organizacji słownictwa w tezaursie, ponieważ te projekty mają taką samą funkcję jak to dzieło (Sosińska-Kalata 2006: 151–152), a lingwistyka komputerowa próbuje odtworzyć to, co stanowiło źródło popularności i siły tezaursów (por. Jarmasz, Szpakowicz 2001: 117). Omawiając nowoczesne projekty leksykograficzne odwołujące się do tradycji tezaurya, trzeba wymienić system WordNet i jego polski odpowiednik, czyli Słowosieć, a także inne tego typu struktury, które zwane są często elektronicznymi tezaurusami języków naturalnych, połączeniem elektronicznych tezaursów z rozbudowaną siecią relacji leksykalno-semantycznych czy po prostu tezaurusami – słownikami pojęciowymi (<http://nlp.pwr.wroc.pl/en/slowosiec-20/opis-projektu-slowosieci-20>; Dobrowoska 2013: 254; Dziob, Łazarewicz 2011: 3). Nie są to jednak jedyne określenia, pojawiają się też inne, takie jak np. słownik semantyczny, sieć semantycznych relacji leksykalnych, baza jednostek leksykalnych uporządkowanych według kryterium semantycznego. W. Kajtoch, opisując różne tradycje warunkujące powstanie polskich słowników niealfabetycznych, pośród nich wskazuje też na nierozwiniętą w Polsce tradycję związaną z *Tezaurusem* Rogeta. Na końcu tych rozważań badacz wymienia nowoczesny sposób budowy tezaursów za pomocą sieci WordNet czy polskiej Słowosieci (Kajtoch 2008: 44–45). Nie można lekceważyć tej rozwijającej się szybko i prężnie dziedziny łączącej w sobie wielowiekową tradycję badań leksykograficznych z nowoczesnymi środkami technologicznymi (np. polska Słowosieć ma już wersję 2.3). Warto jednak podkreślić, że systemy te nie korzystały bezpośrednio z *Tezaurya* Rogeta, ponieważ słownik ten został odrzucony ze względu na problem techniczny, system komputerowy nie był kompatybilny z tezaurusem, nie umożliwiał więc śledzenia automatycznie powiązań między słowami (Jarmasz, Szpakowicz 2001: 117). Jak twierdzili bowiem

badacze, dane z tezaursów tradycyjnych dobre są dla człowieka, a nie dla maszyny, dlatego nie mogą być one dla komputera źródłem informacji o relacjach pomiędzy słowami. Projekty te korzystają natomiast z ontologii rozumianych w znaczeniu informatycznym jako schemat pojęć i łączących je relacji, stanowiący reprezentację pewnego zasobu wiedzy (por. Dobrowolska 2013: 255; oraz <http://pl.wikipedia.org/wiki/Ontologia-termin> „ontologia” w informatyce i podejściu systemowym). Największą formalną i publiczną ontologią istniejącą obecnie jest ontologia Suggested Upper Merged Ontology (SUMO). Została ona wykorzystana w amerykańskim Wordnecie do nadrzędnej organizacji jego struktury, choć, co warto podkreślić, polska Słowność z niej nie korzysta (por. Dobrowolska 2013: 255–256). Okazuje się jednak, że tego typu ontologie poświęcają znacznie mniej uwagi zawłościom związanym ze związkami semantycznymi niż ich hierarchizacji, więc konieczne staje się ich dopracowanie czy dopełnienie (Cassidy 2000).

Badacze upomnieli się więc ponownie o tezaurs Rogeta, który może stanowić dobre uzupełnienie systemów takich jak WordNet. Systemom takim zarzuca się niewielkie połączenia między różnymi częściami mowy i nieuwzględnianie wszystkich rzeczowników, np. nazw własnych. Taka integracja systemów leksykalnych jest więc bardzo pożądana, ponieważ pozwalałaby pokazać lepiej system relacji semantycznych. Umożliwiają ją skomputeryzowane wersje Rogetowskiego *Tezaurusu*, które są tworzone dla potrzeb technologii NLP (natural language processing). Jedną z nich jest elektroniczna wersja *Tezaurusu* z 1911 roku (Project Gutenberg), która stała się podstawą dla *Factotum*, czyli sieci semantycznej (Jarmasz, Szpakowicz 2001: 120). Jej twórca P. Cassidy twierdzi bowiem, że Rogetowski tezaurs przeorganizowany w r. 1911 do hierarchicznie usystematyzowanej sieci semantycznej ma użyteczniejszą strukturę niż poprzednia wersja i posiada alternatywną strukturę do tej zawartej w WordNecie, niemniej jednak konieczny jest ciągły wspólny wysiłek, by mógł być on używany jako użyteczny dla takich projektów jak WordNet. Badacz mówi o konieczności stworzenia jednego standaryzowanego systemu semantycznego, który mógłby być używany w różnych systemach komputerowych (Cassidy 2000).

Kolejną bazą jest komputerowa wersja edycji *Thesaurusu* Rogeta z 1987 roku wydane go przez wydawnictwo Penguin, która jest podstawą *Electronic Lexical Knowledge Base*. Badacze zainspirowani sukcesem WordNetu proponują projekt słownikowej bazy danych stworzonej przez skomputeryzowanie Rogeta, począwszy od systemu klasyfikującego, a skończywszy na indeksie. Jak twierdzą, wielkie bazy elektroniczne jak WordNet poza zaletami mają także swoje wady, które mogłyby być zniwelowane poprzez zintegrowanie wiedzy związanej z leksyką na podstawie Rogeta i połączenie jej z WordNetem. Jedną z najbardziej interesujących cech Rogetowskiego *tezaurusu* są bowiem jego bogate związki semantyczne, powiązania semantyczne, które mogłyby być nazwane z pomocą WordNetu (Jarmasz, Szpakowicz 2001: 117–119 i n. – szczegółowy opis projektu).

Widać więc, że obecnie badacze próbują udoskonalić elektroniczne bazy danych poprzez ich powiązanie z tradycją Rogetowską. Słownik ten ciągle więc jest

bogatym źródłem wiedzy leksykalnej i jego wartość potwierdzana przez 150 lat zyskuje znów nowy wymiar w lingwistyce komputerowej. Choć jest trudny do przełożenia w prosty sposób na automatyczny język maszynowy, okazuje się jednak dobrym źródłem do uzupełnienia opisu bardziej skomplikowanych relacji semantycznych. Na styku nowoczesnej technologii w służbie lingwistyki i badań na temat dzieła z ponad 150-letnią tradycją ciągle powstają ciekawe projekty i prace badawcze odwołujące się jednocześnie do Rogetowskiego tezaurusu i leksykalnych baz danych takich jak WordNet. Przykładem tego mogą być poza wymienionymi powyżej projektami także różnorodne badania porównawcze, np. A. Kilgariffa i C. Yallopa. Zestawiają oni z sobą różne rodzaje tezaurusów, poczynając od tezaurusów w stylu Rogeta stworzonych, żeby pomóc piszącym znajdować synonimy, poprzez bazy WordNet i EuroWordNet, aż po tezaury tworzone ręcznie, by wspomóc system wyszukiwania informacji, oraz tezaury tworzone automatycznie przez przetwarzanie zasobu słownictwa i wyszukiwanie podobieństw między słowami. Badaczy interesuje zagadnienie polisemii w kontekście struktury tezaurusu, dlatego przeprowadzając eksperymenty językowe na podstawie porównania zawartości dzieł słownikowych, próbują odpowiedzieć na pytanie, czy znaczenia bliskie sobie leksykograficznie są także blisko siebie w tezaurusach (Kilgariff, Yallop 2000). Tego typu interesujących badań jest znacznie więcej.

Zakończenie

Jak twierdzą niektórzy, mimo iż George Miller (twórca Wordnetu) stworzył ciekawy system, nie można się oprzeć myśli, że Roget zrobiłby to lepiej, gdyby miał komputer (Jarmasz, Szpakowicz 2001: 117). *Tezaurus* Rogeta to jeden z najlepszych i najbardziej rozpoznawalnych w leksykografii europejskiej słowników onomazjologicznych. Jego recepcja przez 162 lata wskazuje z jednej strony na ciągłość tradycji tak rozumianego tezaurusu, z drugiej strony na zmienność leksykografii europejskiej, która odwołując się do tradycji, ciągle tworzy coś nowego. Mimo głosów krytycznych wobec tej koncepcji wciąż jest ona obecna w leksykografii europejskiej i stanowi w XXI wieku ważne źródło inspiracji. Trzeba mieć nadzieję, że zostanie ona także bardziej zauważona i doceniona w leksykografii polskiej.

Bibliografia

- ANUSIEWICZ Janusz, SKAWIŃSKI Jacek (1996): *Słownik polszczyzny potocznej*. – Warszawa: PWN.
- BURKHANOV Igor (2010): *Lexicography: a Dictionary of Basic Terminology*. – Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.
- CASSIDY Partick J. (2000): An Investigation of the Semantic Relations in the Roget's Thesaurus: Preliminary Results (online: <http://micra.com/factotum/cicparp.doc>, dostęp: 1 października 2014).

- DAKOVIĆ Sybilla (2011): Słowniki tematyczne w Polsce, Serbii i Chorwacji, w: *südslavistik-online*, nr. 3 (April 2011), 33–52 (online: <http://www.suedslavistik-online.de/03/dakovic.pdf> 2011, dostęp: 1 października 2014).
- DOBROWOLSKA Marta (2013): Hiponimy miejsca w polskiej Słowosieci i amerykańskiej bazie WordNet. – *LingVaria* VIII, nr 1 (15), 253–266.
- DZIOB Agnieszka, ŁAZAREWICZ Paulina (2011): Słowosieć jako narzędzie wspomagające pracę tłumacza. – *Rocznik Kognitywistyczny*, V, 33–42.
- ESQUIBEL Joanna (2011): Użyteczność tezaursusa w studium diachronicznym dla językoznawcy historycznego – Historical Thesaurus of the Oxford English Dictionary. – [w:] Włodzimierz GRUSZCZYŃSKI, Laura POLKOWSKA (red.): *Problemy leksykografii. Historia – metodologia – praktyka*, LEXIS. – Warszawa, 143–160.
- GRUSZCZYŃSKI Włodzimierz (2000): *Wokabularze ryskie na tle XVI-i XVII-wiecznej leksykografii polskiej*. – Warszawa: Aspra.
- HÜLLEN Werner (2004): *A History of Roget's Thesaurus. Origin, Developments and Design*. – Oxford: Oxford University Press.
- HÜLLEN Werner (2009): *Networks and Knowledge in Roget's Thesaurus*. – Oxford: Oxford University Press.
- JARMASZ Mario, SZPAKOWICZ Stan (2001): Roget's Thesaurus as an Elektronical Lexical Knowledge Base. – [w:] Włodzimierz GRUSZCZYŃSKI (red.): *Nie bez znaczenia. Prace ofiarowane Profesorowi Zygmuntowi Saloniemu z okazji jubileuszu 1500 dni pracy naukowej*. – Białystok: Wydawnictwo Uniwersytetu w Białymstoku, 117–138.
- KAJTOCH Wojciech (2008): *Językowe obrazy świata i człowieka*, t. I. – Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- KILGARRIFF Adam, YALLOP Colin (2000): What's in a thesaurus?. In Proceedings of Second Conference on Language Resources and Evaluation. Supported by Macquarie University Visiting Researcher Fund., 1371–1379 (online: http://www.researchgate.net/publication/228691171_ITRI-00-28_What%27s_in_a_thesaurus, dostęp 12 października 2014).
- KITA Małgorzata, POLAŃSKI Edward (2002): *Słownik tematyczny języka polskiego*. – Łódź: Literatura.
- KLÉGR Aleš (2000): Rogetuv thesaurus a onomazjologická lexikografie. – *Časopis pro moderní filologii* 82, č. 2, 65–84.
- KLÉGR Aleš (2007): *Tezaurus jazyka českého. Slovník českých slov a frází souznačných, blízkých a příbuzných*. – Praha: NLN.
- KUCAŁA Marian (1957): *Porównawczy słownik trzech wsi małopolskich*. – Wrocław: Zakład Narodowy imienia Ossolińskich.
- MARKOWSKI Andrzej (1990): *Leksyka wspólna różnym odmianom polszczyzny*, t. I, II. – Warszawa: Wiedza o kulturze.
- MIKOŁAJCZAK-MATYJA Nawoja (2005): Hierarchiczne uporządkowanie słownictwa a tradycja leksykografia, *Investigationes Linguisticae*, vol. XII, Poznań, December. (online: http://www.staff.amu.edu.pl/~inveling/pdf/matyja_inve12.pdf, dostęp: 1 października 2014).
- MIODUNKA Władysław (1980): *Teoria pól językowych. Społeczne i indywidualne ich uwarunkowania*. – Kraków: Zeszyty Naukowe Uniwersytetu Jagiellońskiego.
- MIODUNKA Władysław (1989): *Podstawy leksykologii i leksykografii*. – Warszawa: PWN.
- MORAL del Rafael (1998): *Diccionario temático del Español*. – Madrid: Eguilaz.
- PIOTROWSKI Tadeusz (1994): *Z zagadnień leksykografii*. – Warszawa: PWN.
- REMBISZEWSKA Dorota Krystyna (2005): Słownictwo tematyczne w dwudziestowiecznych opracowaniach leksykograficznych. – *Poradnik Językowy*, z. 8, 37–51.

- SIERADZKA-BAZIUR Bożena (2012): O projekcie Słownika pojęciowego języka staropolskiego w kontekście prac dotyczących najdawniejszego słownictwa. – *Język Polski* XCII, 81–91.
- SOSIŃSKA-KALATA Barbara (2006): Tezaury w zmieniającym się środowisku wyszukiwania informacji. – [w:] Barbara SOSIŃSKA-KALATA, Ewa CHUCHRO, Włodzimierz DASZEWSKI (red.): *Informacja w sieci*. – Warszawa, 140–156 (online <http://bbc.uw.edu.pl/Content/3/11.pdf>, dostęp: 2 października 2014).
- URBAŃCZYK Stanisław (1962): Wokabularz trydencki. – *Język Polski* XLIII 21, 15–29.
- ZGÓŁKOWA Halina (2013): *Słownik minimum języka polskiego*. – Kraków: Universitas.
- ŻMIGRODZKI Piotr (2009): *Wprowadzenie do leksykografii polskiej*. – Katowice: Wydawnictwo Uniwersytetu Śląskiego.